

Personal Data Manifesto - Tigistu Korga

Prior to this class, I perceived data as an objective entity awaiting accurate analysis. Over time, I learned that this view was incomplete. Data is not reality itself. It is a representation of reality, shaped by decisions about what to measure, how to measure it, and what gets left out. Because of this, data science is not the extraction of truth from neutral inputs, but a process of reasoning with imperfect and incomplete representations of the world.

This course reshaped my perspective on data by forcing me to confront uncertainty, bias, and interpretation at every stage of analysis. Writing correct code matters, but it is not the core of the work. Judgment is.

I. Data is constructed through human choices

Data does not appear naturally in clean tables. It is created through human systems, tools, and priorities. Choices made during collection and organization shape what a dataset can represent long before analysis begins. While models like the DIKW pyramid suggest a linear progression from data to knowledge, my experience showed that interpretation is present from the start.

This became especially clear when working with data that I helped generate myself, where small decisions about how measurements were taken directly shaped the results. Jill Lepore's file cabinet metaphor further clarified that data is stored and organized according to human logic, not natural order. What gets recorded reflects priorities, convenience, and power. Because data is constructed, I approach every dataset by asking what choices made it possible and what those choices exclude.

II. Data science is an interpretive practice

Computation alone does not produce insight. Tables, statistics, and visualizations require interpretation, and analytical choices shape the meaning that emerges. Decisions about aggregation, scale, filtering, and framing all influence what patterns appear important.

Through repeated analysis and visualization work, I saw how the same data could tell different stories depending on how it was presented. This reinforced the idea emphasized in our discussions of data visualization and data humanism: that charts are not neutral outputs but arguments about what deserves attention, a perspective emphasized in our discussions of data humanism. Because of this, explanation is essential. If an analysis cannot be clearly explained, it is incomplete.

III. Data has limits, and restraint matters

Data is effective at describing patterns, trends, and differences. It is far less effective at explaining causes or motivations on its own. Treating data as capable of answering “why” questions without additional context leads to overconfidence and weak conclusions.

Throughout the semester, I learned to slow down and recognize when the data did not support strong claims. Correlation is not explanation, and clean results do not guarantee meaningful insight. Responsible data work requires restraint and a willingness to acknowledge uncertainty. Sometimes, the most honest conclusion is that the data cannot answer the question being asked.

IV. Data science reflects the analyst

Data science is shaped by the person doing the work. The questions we ask, the problems we prioritize, and the interpretations we find convincing are influenced by our background and values. Pretending to be fully objective hides these influences rather than eliminating them.

Engaging with ideas from human-centered and feminist approaches to data made me more aware of how perspective affects analysis. Acknowledging this does not weaken data science. It makes it more transparent and accountable.

Applying my data science process

If I were starting a new data project, I would begin by examining how the data was collected and what it represents. I would explore structure, missingness, and potential bias before choosing any methods. I would prioritize clarity over complexity and select techniques appropriate to the question rather than those that appear impressive.

At every stage, I would document assumptions and limitations, understanding that data science is a process of reasoning under constraint, not a search for certainty.

Conclusion

To someone who has not taken this class, the most important lesson is that data science is not about producing definitive answers. It is about asking careful questions, making intentional choices, and communicating results honestly. Data is powerful, but only when handled with humility and awareness of its limits.

This class changed how I work with data. I no longer see it as neutral input to be processed, but as a human artifact that demands responsibility, skepticism, and clarity.